

生体模倣型ビジョンシステムによる動画像認識に関する研究

代表研究者 小渡 悟 沖縄大学人文学部福祉文化学科 准教授

目 的

現代社会において労働力やネットワークの端末として様々な機械が普及している。ホンダの ASIMO を始めとする人型ロボットに対する関心は高く、社会で働く機械としてはヒトに近い構造を持つロボットの方が安心を得られているのが分かる。機械（人型ロボット）がヒトの作業を学習したり、あるいは身振り手振りでコミュニケーションを行うとする場合、まず必要になるのがヒトの手の形状および動作の推定である。ロボットの目となるビデオカメラから入力されたヒトの自然な身振り動作を実時間で解析することは自然なコミュニケーションのために非常に重要な要素技術となっている。

本研究では高齢者や障害者への介護や補助に人型ロボットを利用する場合に必要な要素技術のひとつであるヒトと機械のインタフェースとしてのジェスチャ認識について検討する。これはロボットの Sense Plan Action（認識、計画、動作）における Sense の機能に該当するものであり、環境内の対象物体の検出ならびに識別に対応する。

ここでヒトと機械のインタフェースとしてのジェスチャ認識を考えると、手指の運動方向、移動量などの手指の運動認識と手指の形、向きなどの手指の形状認識などがある。運動認識においては高解像度による部分的な詳細情報はあまり必要無く、移動物体の検知など外界の変化に対する情報を収集する機能が必要である。それに対して形状認識においては詳細な情報を得るための機能が必要である。しかし、現在のコンピュータビジョンシステムの能力は、ヒトの視覚と比べると柔軟性、頑健性、適応性などの点でまだ劣っている。

ヒトは興味のある部分について詳細な情報を収集する中心視と移動物体の検知など外界の変化に対する情報を収集する周辺視といった視覚機能を使い分け、効率よく処理を行っている。このように中心の解像度が高く中心から周辺へと離れていくに従って解像度が低くなる構造を持つ中心窩視覚は対数極座標変換（Log Polar Mapping; 以下 LPM と略す）によって表現することができる。また猫の視覚皮質のニューロン群で起こる同期発火現象を説明するために提案されたニューラルネットワークモデルとして、Pulse-Coupled Neural Network（以下 PCNN と略す）がある。このモデルは画像情報のある時系列のデータに分散して埋め込むことができる特徴を有しており、画像の特徴がある一定周期内に分散されることで、画像中に存在する物体のセグメンテーションおよびエッジ抽出をすることが可能となる。

生物の視覚が現在のコンピュータビジョンシステムとはかなり異なった仕組みに基づき高い処理能力を発揮していることから、この生物の視覚系の機能を模倣することで実環境でのコンピュータビジョンシステムの能力を向上させることが期待できる。

本研究のシステムでは生体の視覚系の知見をコンピュータによる物体認識処理に応用することで実環境でも安定して利用できるビジョンシステムについて検討を行う。その有効性をビデオカメラによる動画像からの移動物体の位置と形状の推定を用いたヒトと計算機とのインタフェースに適用することで示す。

方 法

本研究では非接触型デバイスであるビデオカメラを用い、画像からユーザの手の位置と形状を推定することで、ユーザに対して拘束感や不自然な使用感を与えないヒトと計算機のインタフェースシステムを検討した。

本システムでは、ビデオカメラからの取り込み画像の前処理として改良型 PCNN を適用することで動画像中の物体の正確な位置抽出、およびエッジ等の抽出を行う。改良前の PCNN モデルは画像情報のある時系列のデータに分散して埋め込むことができる特徴を有しており、画像の特徴がある一定周期内に分散されることで、画像中に存在する物体のセ

グメンテーションおよびエッジ抽出をすることが可能だが静止画像を対象としたモデルである。そのため本研究では PCNN の入力を動画像のフレーム方向（時間方向）に拡張することで、移動物体の抽出、およびエッジ等の特徴抽出を行う。

抽出された対象物を中心として LPM を用いて直交座標系から対数極座標系に変換を行う。この LPM による処理の利点は、少ない画素数で高解像度、広域の作業範囲を得るだけでなく、スケーリング不変性、変換中心に対する回転不変性を持っており、また、画像のデータ量を減らし画像処理の計算時間を減少させることが期待できる。抽出した領域から高次局所自己相関特徴を算出し、学習ベクトル量子化を用いたニューラルネットワークにて手形状の識別を行なう。この機能を用いてポインティング操作を実現する。手領域の位置の軌跡を用いジェスチャ認識を行うことで、手を用いたポインティング操作とジェスチャ操作という二種類の操作を連携して利用することを実現する。ここで、計算機の入力デバイスとしての手指動作によるマウス様な機能の実現のためにジェスチャ認識を用いることを考慮すると、任意の動きからジェスチャ認識を行なうよりも、ユーザが入力操作を意図したときにジェスチャを認識したほうが誤動作を招かなくて都合がよい。そこで本システムでは、入力キーとなる手形状をカメラの前で行なうことで、マウス様な機能のオン/オフを行なうことができるようにした。またポインティング操作とジェスチャ操作の切り替え(モードセレクト)は、あらかじめ操作の切り替え用の手形状を決めておき、その手形状を提示することで行う。

結 果

本研究のシステムでは生体の視覚系の知見をコンピュータによる物体認識処理に応用することで実環境でも安定して利用できるビジョンシステムについて検討を行った。ビデオカメラから取り込んだ画像に対して PCNN の入力を動画像のフレーム方向（時間方向）に拡張(改良型 PCNN)することで、移動物体の抽出、およびエッジ等の特徴抽出を行った。改良型 PCNN で対象領域を決定後、対象領域を中心とした LPM により LPM 画像に変換することで、スケーリング不変性、変換中心に対する回転不変性を得ることができ、LPM 画像から平行移動に対して不変な高次局所自己相関特徴を抽出することで、入力画像上の対象の大きさの変動や対象の回転に対して頑健な特徴量を得ることができた。また、より快適な計算機とのインタラクションを実現するために、ポインティングによる直接操作型の機能に加え、手の動作軌跡からのシンボリックサインを識別する機能を組み込んだ。評価実験にて、通常の計算機とデジタルビデオカメラにて構築したシステムを用いて実用的な速度で安定して検出することができ、また手領域の追跡、抽出されたフレーム区間に対してシンボリックサインの推定が実現した。提案手法は従来手法と比較して、対象物体の細かな形状や動作が扱え、しかも、大きさの変動や対象の回転に対して頑健な特徴量を与えるなどの優位性を示した。今後は、認識部の写像関係モデル、高次局所自己相関特徴、改良型 PCNN の入力部などは、画像中の局所領域での処理となるため、各段階での処理の並列化が可能であることから、本研究でのビジョンシステムは並列処理による高速化、およびチップ化による高速化が比較的容易に実現できると思われる。

研究発表

- 1) Satoru Odo, "Sequential Image Analysis using Pulse Coupled Neural Networks for Pre-processing," Intelligent Engineering Systems Through Artificial Neural Networks, vol. 15, p.p.597 - 606, 2005.
- 2) 星野聖, 小渡悟, "画像に含まれる身体形状を判定する方法及び非接触型ポインティング・デバイスの実現方法," 特開 2005 - 122492 号, 2005.

連絡先

沖縄県那覇市字国場 555 番地 沖縄大学人文学部 小渡悟
E-mail: odo@okinawa-u.ac.jp